ECE-374-B: Lecture 0 - Logistics and Strings/Languages

Instructor: Abhishek Kumar Umrawal January 16, 2024

University of Illinois at Urbana-Champaign

Course Administration

Instructional Staff

- Instructor:
 - Abhishek Kumar Umrawal
- Teaching Assistants:
 - Anthea (Anxue) Chen
 - Jack (Ziheng) Chen
 - Sung Woo Jeon
 - Hongyu Shen

- Sumedh Vemuganti
- Weiyang Wang
- Haoyuan You
- Hongbo Zheng
- Office hours: TBD, See course webpage
- Contacting us: Use <u>private notes</u> on Piazza to reach course staff. Direct email only for sensitive or confidential information.

About your instructor – Basic info

- Name: Abhishek Kumar Umrawal
- Webpage: ece.illinois.edu/about/directory/faculty/aumrawal
- Email: aumrawal@illinois.edu
- Office: ECEB 3054
- Office hours: TBD

About your instructor – Education

- Purdue University, Ph.D. in Industrial Engineering Dissertation: Machine Learning Algorithms for Influence Maximization on Social Networks
- Purdue University, MS in Economics
- Indian Institute of Technology (IIT) Kanpur, MS in Statistics

About your instructor – Prior teaching experience

• University of Maryland, Visiting Lecturer of Computer Science and Electrical Engineering

About your instructor – Research interests

Core areas:

- 1. Combinatorial optimization
- 2. Approximation algorithms
- 3. Statistical learning theory
- 4. Reinforcement learning (RL)
- 5. Causal inference

About your instructor – Research interests

Core areas:

- 1. Combinatorial optimization
- 2. Approximation algorithms
- 3. Statistical learning theory
- 4. Reinforcement learning (RL)
- 5. Causal inference

Applications:

- 1. Social networks
- 2. Promotional marketing
- 3. Intelligent transportation
- 4. Product recommendation

About your instructor – Research interests

Core areas:

- 1. Combinatorial optimization
- 2. Approximation algorithms
- 3. Statistical learning theory
- 4. Reinforcement learning (RL)
- 5. Causal inference

Applications:

- 1. Social networks
- 2. Promotional marketing
- 3. Intelligent transportation
- 4. Product recommendation





If you are interested in working with me then please send me an email with subject line 'Expressing interest in working with you on research' with a brief description of your interests and skills with no attachments. Please do so only after the second midterm so that you have spent enough time learning algorithms. If you are interested in working with me then please send me an email with subject line 'Expressing interest in working with you on research' with a brief description of your interests and skills with no attachments. Please do so only after the second midterm so that you have spent enough time learning algorithms.

Preferred (but not required) skills:

- Mathematical thinking
- Probability and statistics
- Python programming graphs, object-oriented programming, recursion, etc.
- Algorithms (you're doing it this semester!)

You may fill out this form to provide further information.

This semester, the two sections will be run completely independently.

- Different lectures.
- Different homeworks, quizzes, exams.
- Different grading policies.

This semester, the two sections will be run completely independently.

- Different lectures.
- Different homeworks, quizzes, exams.
- Different grading policies.

Section B will be in-person only. Recordings will be attempted but not guaranteed.

Online resources

 Webpage: General information, announcements, homeworks, quizzes, course policies will be available at https://ecealgo.com.

• Submission(Gradescope): Written homework submission and grading, regrade requests. Exams wil be uploaded there as well.

- <u>Communication(Piazza)</u>: Announcements, online questions and discussion, contacting course staff (via private notes).
- Gradebook (Canvas): Announcements, online questions and discussion, contacting course staff (via private notes).

See course webpage for links.

Important: Check Piazza/course web page at least once each day.

Discussion Sessions/Labs

- 50min problem solving session led by TAs.
- Two times a week.
- Go to your assigned discussion section.
- Bring pen and paper!

Discussion Sessions/Labs

- 50min problem solving session led by TAs.
- Two times a week.
- Go to your assigned discussion section.
- Bring pen and paper!

Discussion sections will have questions that appear on the homework. If, you skip, you're just making more work for yourself later.

Again all policy information should be on course website: https://ecealgo.com.

Any questions?

Over-arching course questions

This course introduces three distinct fields of computer science research:

- Computational complexity.
 - Given infinite time and a certain machine, is it possible to solve a given problem.
- Algorithms.
 - Given a deterministic Turing machine, how fast can we solve certain problems.
- Limits of computation.
 - Are there tasks that our computers cannot do and how do we identify these problems?

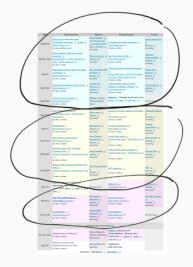
When someone asks you, "How fast can you compute problem X", they are actually asking:

- Is X solvable using the deterministic Turing machines we have at our disposal?
- If it is solvable, <u>can we find the solution efficiently</u> (in poly-time)?
- If it is solvable but we don't have a poly time solution, what problem(s) is it most similar to ?

Course Structure

Course divided into three parts:

- Basic automata theory: finite state machines, regular languages, hint of context free languages/grammars, Turing Machines.
- Algorithms and algorithm design techniques.
- Undecidability and NP-Completeness, reductions to prove intractability of problems.





Algorithmic thinking.



- Learn/remember some basic tricks, algorithms, problems, ideas.
- Understand/appreciate limits of computation (intractability).
- Appreciate the importance of algorithms in computer science and beyond (engineering, mathematics, natural sciences, social sciences, ...).

Formal languages and complexity (The Blue Weeks!)

First 5 weeks devoted to language theory.

First $\frac{5}{5}$ weeks devoted to language theory.

But why study languages?

Consider the following problem:

Problem Given two *n*-digit numbers *x* and *y*, compute their product.

Grade School Multiplication

Compute "partial product" by multiplying each digit of y with x and adding the partial products.

 $3141 \\
 \times 2718 \\
 25128 \\
 3141 \\
 21987 \\
 <u>6282 \\
 8537238$ </u>

Time analysis of grade school multiplication

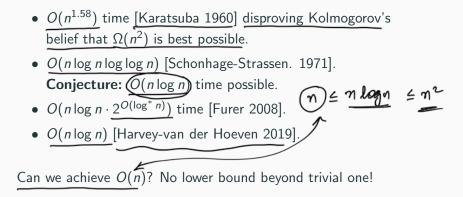
- Each partial product: $\Theta(n)$ time
- Number of partial products: $\leq n$
- Adding partial products: n additions each $\Theta(n)$ (Why?)
- Total time: $\Theta(n^2)$
- Is there a faster way?

$$f(n) is (a) if f(n) = cn \qquad for some c \ge c$$

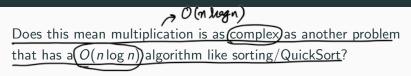
$$|| || O(n) || || \le cn \qquad for some c \ge c$$

$$|| || \Omega(n) || || \ge cn \qquad \dots$$

Fast Multiplication



Equivalent Complexity

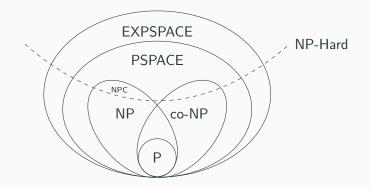


Does this mean multiplication is as complex as another problem that has a $O(n \log n)$ algorithm like sorting/QuickSort?

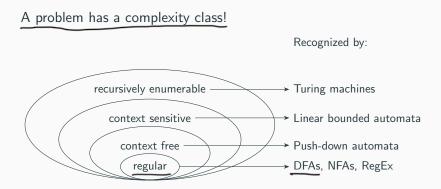
How do we compare? The two problems have:

- Different inputs (two numbers vs n-element array).
- Different outputs (a number vs n-element array).
- Different entropy characteristics (from a information theory perspective).

An algorithm has a runtime complexity.



Languages, Problems and Algorithms ... oh my! III



Problems do not have run-time since a problem \neq the algorithm used to solve it. *Complexity classes are defined differently.* How do we compare problems? What if we just want to know if a problem is "computable".

Algorithms, Problems and Languages ... oh my! I

Definition

- 1. An algorithm is a step-by-step way to solve a problem.
- A problem is some question that we'd like answered given some input. It should be a decision problem of the form "Does a given input fulfill property X."
- A Language is a set of strings. Given a alphabet, ∑ a language is a subset of Σ*.

Definition

- 1. An algorithm is a step-by-step way to solve a problem.
- A problem is some question that we'd like answered given some input. It should be a decision problem of the form "Does a given input fulfill property X."
- 3. A Language is a set of strings. Given a alphabet, Σ a language is a subset of Σ^* . A language is a formal realization of this problem. For problem X, the corresponding language is:

$$\begin{split} L &= \{w - w \text{ is the encoding of an input y to problem X and} \\ \text{the answer to input y for a problem X is "YES"} \\ \hline A \text{ decision problem X is "YES" is the string is in the language.} \end{split}$$

How do we define the multiplication problem as a language?

Define L as language where inputs are separated by comma and output is separated by —.

Machine accepts a $\overline{x^*y=z}$ if $\overline{x^*y=z''}$ is in L. Rejects otherwise. z*y=z 1*z=2 V Machine YES 1*3=4 X U NO \vdots NO How do we define the multiplication problem as a language?

Define L as language where inputs are separated by comma and output is separated by —.

Machine accepts a $x^*y=z$ if " $x^*y=z$ " is in L. Rejects otherwise.

$$\underline{L_{MULT2}} = \begin{cases} \frac{1 \times 1|1}{2 \times 1|2}, & \frac{1 \times 2|2}{2 \times 2|4}, & \frac{1 \times 3|3}{2 \times 3|6}, \dots \\ \vdots & \vdots & \vdots \\ n \times 1|n, & \underline{n \times 2|2n}, & \underline{n \times 3|3n}, \dots \end{cases}$$
(1)

We do the same thing for sorting.

Define L as language where inputs are separated by comma and output is separated by —.

Machine accepts a $[i_1, i_2, ...] = sort(\{i_1, i_2, ...\})$ if "x[]—z[]" is in L. Rejects otherwise.

We do the same thing for sorting.

Define L as language where inputs are separated by comma and output is separated by —.

Machine accepts a $[i_1, i_2, ...] = sort(\{i_1, i_2, ...\})$ if "x[]—z[]" is in L. Rejects otherwise.

$$\underline{L_{Sort2}} = \begin{cases} \underline{1,1} | \underline{1,1} & 1,2 | 1,2 & 1,3 | 1,3,\dots \\ \underline{2,1} | \underline{1,2} & 2,2 | 2,2, & 2,3 | 2,3,\dots \\ \vdots & \vdots & \vdots \\ \underline{0}, 1 | 1,n, & n,2 | 2,n, & n,3 | 3,n,\dots \end{cases}$$
(2)

2,1 2,1 & LSort 2

We do the same thing for sorting.

Define L as language where inputs are separated by comma and output is separated by —.

Machine accepts a $[i_1, i_2, ...] = sort(\{i_1, i_2, ...\})$ if "x[]—z[]" is in L. Rejects otherwise.

$$L_{Sort2} = \begin{cases} 1, 1|1, 1 & 1, 2|1, 2 & 1, 3|1, 3, \dots \\ 2, 1|1, 2, & 2, 2|2, 2, & 2, 3|2, 3, \dots \\ \vdots & \vdots & \vdots \\ n, 1|1, n, & n, 2|2, n, & n, 3|3, n, \dots \end{cases}$$
(2)

If the same type of machine can recognize both languages, then that gives us an upperbound top their hardness.

How do we formulate languages?

Strings

Alphabet

An alphabet is a finite set of symbols.

Examples of alphabets:

- $\Sigma = \{0, 1\}$,
- $\Sigma = \{a, b, c, \ldots, z\},$
- ASCII.
- UTF8.
- $\Sigma = \{\langle (w) \text{forward} \rangle, \langle (a) \text{strafe left} \rangle, \langle (s) \text{back} \rangle, \langle (d) \text{strafe right} \rangle \}$

String Definition

Definition

$$|\epsilon|=0$$

 $|\xi\epsilon_{f}|=1$ ϵ : Empty string

- 1. A string/word over Σ is a finite sequence of symbols over Σ . For example, '0101001', 'string', '(moveback)(rotate90)'
- 2. $x \cdot y \equiv xy$ is the concatenation of two strings
- 3. The length of a string <u>w</u> (denoted by |w|) is the number of symbols in w. For example, |101| = 3, $|\epsilon| = 0$
- 4. For integer $\underline{n \ge 0}$, $\underline{\sum^n}$ is set of all strings over $\underline{\sum}$ of length \underline{n} . $\underline{\sum^*}$ is the set of all strings over $\underline{\sum}$.
- 5. Σ^* set of all strings of all lengths including empty string.

Question: What is $\{0', 1'\}^{*}$? $\{0, 1\}^{*} = \{0, 0, 0, 00, 00, ...\}$

- ϵ is a string containing no symbols. It is not a set
- {e} is a set containing one string: the empty string. It is a set, not a string.

Question: What is $\{\emptyset\}$?

 $|\{p\}| = |$

Concatenation and properties



- If x and y are strings then xy denotes their concatenation.
- Concatenation defined recursively :

•
$$\underline{xy = y}$$
 if $\underline{x = \epsilon}$

•
$$xy = a(wy)$$
 if $x = aw$

- xy sometimes written as x•y.
- concatenation is associative: (uv)w = u(vw) hence write $uvw \equiv (uv)w = u(vw)$
- not commutative: uv not necessarily equal to vu
- The identity element is the empty string ϵ :

$$\epsilon u = u\epsilon = u.$$

Definition v is substring of $w \iff$ there exist strings x, y such that w = xvy.

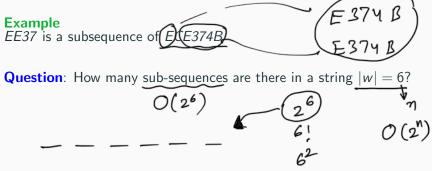
• If
$$x = \epsilon$$
 then v is a prefix of w

• If $y = \epsilon$ then v is a suffix of w

A <u>subsequence</u> of a string w[1...n] is either a subsequence of w[2...n] or w[1] followed by a subsequence of w[2...n].

Example *EE37* is a subsequence of *ECE374B*

A subsequence of a string w[1...n] is either a subsequence of w[2...n] or w[1] followed by a subsequence of w[2...n].



Definition If w is a string then w^n is defined inductively as follows: $w^n = \epsilon$ if n = 0 $w^n = ww^{n-1}$ if n > 0

Question: $(ha)^3 = hahaha$

$$(ha)^{o} = E$$

Answer the following questions taking $\Sigma = \{0, 1\}$. 1. What is Σ^0 ? $\{e\}$ 2. How many elements are there in Σ^n ? $\Sigma^1 = \{o, 1\}$ 3. If |u| = 2 and |v| = 3 then what is $|u \cdot v|$? 2+3=54. Let u be an arbitrary string in Σ^* . What is ϵu ? What is $u\epsilon$? Pattern)

$$\begin{pmatrix}
\varepsilon u = u \\
u \varepsilon = u
\end{pmatrix}$$

Languages

Definition A language *L* is a set of strings over Σ . In other words $L \subseteq \Sigma^*$.

Definition

A language L is a set of strings over Σ . In other words $L \subseteq \Sigma^*$.

Standard set operations apply to languages.

- For languages A, B the concatenation of A, B is
 AB = {xy | x ∈ A, y ∈ B}.
- For languages A, B, their union is $A \cup B$, intersection is $A \cap B$, and difference is $A \setminus B$ (also written as A B).
- For language $A \subseteq \Sigma^*$ the complement of A is $\overline{A} = \Sigma^* \setminus A$.

Definition

Given two sets X and Y of strings (over some common alphabet

 Σ) the concatenation of X and Y is

$$XY = \{xy \mid x \in X, y \in Y\}$$
(3)

Question: $X = \{ECE, CS, \}, Y = \{340, 374\} \implies XY = .$



Σ^* and languages

Definition

1. Σ^{n} is the set of all strings of length *n*. Defined inductively: $\Sigma^{n} = \{\epsilon\}$ if n = 0 $\Sigma^{n} = \Sigma\Sigma^{n-1}$ if n > 02. $\Sigma^{*} = \bigcup_{n \ge 0} \Sigma^{n}$ is the set of all finite length strings 3. $\Sigma^{+} = \bigcup_{n \ge 1} \Sigma^{n}$ is the set of non-empty strings.

Definition

A language L is a set of strings over Σ . In other words $L \subseteq \Sigma^*$.

Question: Does
$$\Sigma^*$$
 have strings of infinite length? $MO!$
 $\Sigma^n = \Sigma' \Sigma^{n-1} \qquad \Sigma^+ = \Sigma^n \setminus \{e\}$

Problem

Consider languages over $\Sigma = \{0, 1\}$.

- 1. What is \emptyset^0 ?
- 2. If |L| = 2, then what is $|L^4|$?
- 3. What is \emptyset^* , $\{\epsilon\}^*$?
- 4. For what L is L* finite?
- 5. What is \emptyset^+ ?
- 6. What is $\{\epsilon\}^+$?



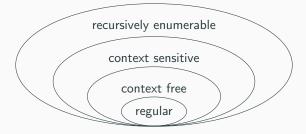
Let's review what we learned.

- A <u>character</u>(*a*, *b*, *c*, *x*) is <u>a unit of information represented by a</u> <u>symbol</u>: (letters, digits, whitespace)
- A alphabet(Σ) is a set of characters
- A <u>string(w</u>) is a <u>sequence of characters</u>
- A language(A, B, C, L) is a set of strings

Let's review what we learned.

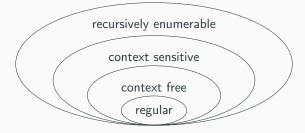
- A character(*a*, *b*, *c*, *x*) is a unit of information represented by a symbol: (letters, digits, whitespace)
- A alphabet(Σ) is a set of characters
- A string(w) is a sequence of characters
- A language(A, B, C, L) is a set of strings
- A grammar(G) is a set of rules that defines the strings that belong to a language

Languages: easiest, easy, hard, really hard, reallyⁿ hard



- Regular languages.
 - Regular expressions.
 - DFA: Deterministic finite automata.
 - NFA: Non-deterministic finite automata.
 - Languages that are not regular.
- Context free languages (stack).
- Turing machines: Decidable languages.
- TM Undecidable/unrecognizable languages (halting theorem). 38

Languages: easiest, easy, hard, really hard, reallyⁿ hard



- Regular languages.
 - Regular expressions. (- Next lecture)
 - DFA: Deterministic finite automata.
 - NFA: Non-deterministic finite automata.
 - Languages that are not regular.
- Context free languages (stack).
- Turing machines: Decidable languages.
- TM Undecidable/unrecognizable languages (halting theorem). 38

Check the course website (https://ecealgo.com) for lab and hw schedule.